IRTG 1792 Discussion Paper 2019-019

Modelling Systemic Risk Using Neural Network Quantile Regression

Georg Keilbar [*]
Weining Wang [*] [*2]

[*] Humboldt-Universität zu Berlin, Germany
[*2] University of London, United Kingdom

This research was supported by the Deutsche Forschungsgesellschaft through the International Research Training Group 1792 "High Dimensional Nonstationary Time Series".

http://irtg1792.hu-berlin.de
ISSN 2568-5619

Noname manuscript No.
(will be inserted by the editor)

Modelling Systemic Risk Using Neural Network Quantile Regression

Georg Keilbar · Weining Wang

Received: date / Accepted: date

Abstract We propose an approach to calibrate the conditional value-at-risk (CoVaR) of financial institutions based on neural network quantile regression. Building on the estimation results we model systemic risk spillover effects across banks by considering the marginal effects of the quantile regression procedure. We adopt a dropout regularization procedure to remedy the well-known issue of overfitting for neural networks, and we provide empirical evidence for the favorable out-of-sample performance of a regularized neural network. We then propose three measures for systemic risk from our fitted results. We find that systemic risk increases sharply during the height of the financial crisis in 2008 and again after a short period of easing in 2011 and 2015. Our approach also allows identifying systemically relevant firms during the financial crisis.

Keywords Systemic risk · CoVaR · Quantile regression · Neural networks

Financial support from the Deutsche Forschungsgemeinschaft via the IRTG 1792 "High Dimensional Non Stationary Time Series", Humboldt-Universität zu Berlin, is gratefully acknowledged.

Georg Keilbar
IRTG 1792 "High Dimensional Nonstationary Time Series", School of Business and Economics, Humboldt-University of Berlin, Unter den Linden 6, 10099 Berlin, Germany
E-mail: georg.keilbar@hu-berlin.de

Weining Wang
Chair of Statistics, School of Business and Economics, Humboldt-University of Berlin, Unter den Linden 6, 10099 Berlin, Germany
City, University of London, Northampton Square, London, EC1V 0HB, United Kingdom
E-mail: wangwein@hu-berlin.de

1 Introduction

The issue of systemic risk attracts a lot of attention from academics as well as from regulators in the aftermath of the financial crisis of 2007-2009. Systemic risk refers to banks and other economic agents with substantial importance to the financial system due to their size (*too big to fail*) or their centrality within the financial network (*too interconnected to fail*). A bankruptcy of a systemically important financial institution can lead to the malfunctioning of the financial system or central banks and governments might be under pressure to interfere by bailing out respective firm. Due to these negative externalities, it is a crucial task for central banks and supervising agencies to identify systemically relevant firms.

A conventional quantitative risk measure is value-at-risk (VaR), which measures maximum losses at a certain confidence level. The Basel II Accord introduced VaR as a preferred measure for market risk. However, VaR is not suitable for capturing systemic risk adequately, as it is not capable to analyze the interdependency among firms. Given the subprime mortgage crisis in 2008, the Basel Committee on Banking Supervision has revised its Accords to focus on strong governance and risk management. Basel III is thus set up to control the systemic risk of the whole financial system, and it enforces additional requirements for identifying systemic risk important banks and generates demands on evaluating the interdependency of risk among banks. Many methods for the quantification of systemic risk are proposed. Adrian and Brunnermeier (2016) come up with conditional value-at-risk (CoVaR), a systemic extension of VaR. However, their original approach is restricted to analyze systemic risk in a bivariate context. Namely, they focus primarily on the risk contribution of an individual financial firm to the entire system, controlling for variables indicating general macroeconomic conditions. Hautsch et al. (2014) modify the estimation of CoVaR further to analyze systemic risk in a multiple equation setup using the LASSO. Härdle et al. (2016) follow up this setup, and extend it to a nonlinear regression setting. In the meanwhile, there are numerous other methods for calibrating systemic risk. Acharya et al. (2017) build an economic model of systemic risk and measure the systemic risk externality of a financial institution by the systemic expected shortfall. Brownlees et al. (2012) develop a systemic risk measure capturing the capital shortage given its degree of leverage and marginal expected shortfall. Diebold and Yılmaz (2014) analyze the connectedness of financial firms in a network context using forecast variance decompositions in a vector autoregressive framework.

Nonlinearity is an important issue for the prediction performance of risk measures due to the complex dependency channels of financial institutions (Chao et al. (2015)). Neural networks have proved to be a suitable method for fitting nonlinear functions. This paper provides a new perspective for estimating CoVaR using neural networks. Over the last years, neural networks have become state of the art models for prediction. They have been applied extensively and successfully to various fields, including image classification (Simonyan and Zisserman (2014)) as well as speech recognition problems (Graves

et al. (2013)). We take the off-shelf neural network methodology and apply it to quantify financial risk. Our findings show that the quantile neural network based approach provides a unique angle compared to the linear model for calibrating the systemic risk due to its flexibility. In particular, we find better out-of-sample prediction with our fine-tuned nonlinear neural network relative to the baseline linear model.

There is a big literature on econometrics analysis using neutral network. White (1988) starts to investigate the usefulness of adopting a neural network for economic prediction. Unfortunately, the message is that even with simple neural networks the prediction performance is not ideal due to the overfitting issues. Kuan and White (1994) provide a further overview of neural networks with some basic concepts and theory. White (1992) provides the theoretical foundations of a nonparametric quantile neural network approach allowing for cases of dependent data. In terms of economic risk prediction, Taylor (2000) is concerned with predicting conditional volatility by adopting a quantile neural network approach. Xu et al. (2016) consider a quantile neural network procedure for evaluating VaR in the stock market. Cannon (2011) focuses on the computational perspective of a quantile neural network. We follow the existing approach and take it further for evaluating and forecasting systemic risk.

We briefly summarize the steps of calibrating the systemic risk using a quantile neural network procedure. In the first step, we estimate the VaR for each global systemically important financial institution (G-SIB) from the United States by regressing their stock returns on a set of risk factors using linear quantile regression. Next, we estimate the CoVaRs of the same firms using neural network quantile regression. To characterize the interdependency among banks, we regress the return of one asset on the remaining returns respectively and aggregate the results into a systemic fit. By approximating the conditional quantile with a neural network we aim for capturing possible nonlinear effects. To estimate risk spillover effects across banks we calculate the marginal effects by taking the derivative of the fitted quantile with respect to the other banks' stock returns, evaluated at their VaR. By doing so we come up with a network of spillover effects represented by an adjacency matrix. This adjacency matrix is time-varying, i.e. we estimate a network for each window in our moving window estimation procedure.

In the final step, we propose three systemic risk measures building on the previous results. As a first measure, we propose the *Systemic Fragility Index*, which identifies the most vulnerable banks in a given financial risk network. The second measure is the *Systemic Hazard Index*, which identifies the financial institutions which potentially pose the largest risk to the financial system. These two measures characterize the firm-specific aspects of systemic risk. Thus we propose a third measure which estimates the total level of systemic risk, the *Systemic Network Risk Index*. Our main empirical findings show that systemic risk increased sharply during the height of the financial crisis after the bankruptcy of Lehman Brothers in 2008. Systemic risk remains stable over the last years with two minor spikes in 2011 and 2015. We compare our results to the aggregated SRISK measure of Brownlees and Engle (2016) and find

a strong co-movement of both indices. We also identify systemically relevant financial institutions during the financial crisis. In particular, we identify a risk cluster of four banks, which corresponds to the list of banks that received the largest funding in the course of the bank bailout of 2008. Finally, we compare the predictive performance of our neural network model to a linear baseline model. An out-of-sample prediction comparison shows the superiority of our approach and leads to the conclusion that non-linear effects are crucial for the modelling of systemic risk.

The remainder of this paper is organized as follows. Section 2 provides a brief introduction to neural networks in general and neural network quantile regression in particular. Section 3 describes in detail the methodology of this paper. After establishing the research framework step by step, we present the results in section 4. Section 5 discusses the results and concludes.

2 Neural Network Quantile Regression

2.1 Neural Network Sieve Estimation

Neural networks attract increasing attention due to their success in a variety of prediction problems. Often described as a black box, single hidden layer neural networks can be seen as a special case of the nonparametric sieve estimator, see Grenander (1981) and Chen (2007). With increasing sample size n the complexity of the estimator of h_θ is required to increase appropriately fast. The structure of the neural network sieve is as follows, with $t = 1, 2, \cdots, n$,

$$\begin{aligned} Y_t &= h_\theta(X_t) + \varepsilon_t \\ &= \sum_{m=1}^{M_n} w_m^o \psi \left(\sum_{k=1}^{K} w_{k,m}^h X_{k,t} + b_m^h \right) + b_o + \varepsilon_t \end{aligned} \tag{1}$$

where Y_t is the dependent variable, X_t is a K-dimensional vector of independent variables and ε_t is an error term. The nonlinear activation function $\psi(\cdot)$ is assumed to be fixed and known. Typical choices are sigmoid functions, e.g. $\psi(z) = \tanh(z)$ or the ReLU (rectifier linear unit) function, $\psi(z) = \max(z, 0)$. There are two types of parameters, hidden layer parameters $w_{k,m}^h$ and b_m^h and output layer parameters w_m^o and b^o. The sieve parameter space Θ_n expands with n. In particular, the number of basis functions (i.e. the number of hidden nodes) goes to infinity, $M_n \to \infty$ as $n \to \infty$. Single layer neural networks have proved to be universal function approximators, as shown by Cybenko (1989) for sigmoid activation functions and Hornik et al. (1989) for the general case of bounded, non-constant activation functions. Sonoda and Murata (2017) extend the universal approximation property to unbounded activation functions, which includes the popular ReLU function.

The large sample properties of neural networks have been studied extensively in the literature. Notably, Chen and White (1999) show consistency and asymptotic normality of the nonparametric neural network sieve estimator

under certain regularity conditions. Given that the number of basis functions grows appropriately with increasing sample size, the root mean square convergence rate to an unknown (suitably smooth) true function is of order $o_p(n^{-1/4})$. This rate is crucial to obtain root-n asymptotic normality for plug-in estimators (Chen and Shen (1998)).

All of the above results concern with neural networks with a single hidden layer. The approximation theory and the asymptotic results of deep neural networks, i.e. neural networks with more than one hidden layer, is less understood compared to the shallow neural network case. Johnson (2018) shows that deep neural networks with limited width are not universal function approximators. Rolnick and Tegmark (2017) prove that deep neural networks can learn polynomial functions more efficiently (in terms of number of nodes required) than shallow ones.

2.2 Neural Network Sieves and Quantile Regression

Predominantly, neural networks have been applied to classification and mean regression problems. However, an extension to a quantile regression setting is straightforward. Consider the linear quantile regression equation for a fixed quantile level τ, as formulated in Koenker and Bassett Jr (1978) and Koenker and Bassett Jr (1982).

$$Y_t = X_t\beta + \varepsilon_t, \quad t = 1, \ldots, n \tag{2}$$

with $Q^\tau(\varepsilon_t|X_t) = 0$. In this setting the dependent variable Y_t is modelled as a linear function of independent variables X_t. The linear quantile estimator is then the solution to the following minimization problem:

$$\min_{\beta} \sum_{t=1}^{n} \rho_\tau \left(Y_t - X_t\beta\right) \tag{3}$$

where $\rho_\tau(z) = |z| \cdot |\tau - \mathbf{I}(z < 0)|$ is the quantile loss function. This minimization problem can be formulated as a linear program and can thus be solved by simplex or interior point algorithms. Neural network quantile regression is a nonlinear generalization of this regression framework. Instead of using a linear function, the conditional quantile is approximated by a neural network sieve estimator as defined in 2.1. The resulting optimization problem is nonconvex and cannot be solved by linear programming methods:

$$\min_{\theta} \sum_{t=1}^{n} \rho_\tau \left(Y_t - h_\theta(X_t)\right) \tag{4}$$

A possible alternative is to use the gradient-based backpropagation algorithm of Rumelhart et al. (1988). The asymptotic properties of nonparametric neural network estimators for the conditional quantile are analyzed in White (1992). Under certain regularity conditions the estimator is consistent, see Appendix A. This result holds both for i.i.d. and dependent data.

2.3 Regularization Methods for Model Fitting

Neural networks are prone to overfitting due to their high capacity. An effective tool to counteract overfitting lies in the choice of the structure and the hyperparameters of the neural network. In our single hidden layer setting, the most important hyperparameter is the number of hidden nodes, M_n. Other relevant parameters are the number of epochs and the specification of the learning algorithm. Typically, hyperparameters are selected according to a cross-validation criterion. A different approach is to put an extra penalty term on the weight parameters, $w^h_{k,m}$ and w^o_m. We are considering both L_1 and L_2 penalties which we summarize under the term elastic net (Zou and Hastie (2005)). This penalization method leads to the following optimization problem:

$$\min_{h_\theta} \sum_{t=1}^{n} \rho_\tau \left\{ Y_t - h_\theta(X_t) \right\} + \lambda \left\{ \alpha \| (w^{h\top}_{k,m}, w^{o\top}_m)^\top \|_1 + (1-\alpha) \| (w^{h\top}_{k,m}, w^{o\top}_m)^\top \|_2^2 \right\} \tag{5}$$

where $\| \cdot \|_1$ is the L_1-norm, $\| \cdot \|_2$ is the L_2-norm and $\alpha \in [0,1]$ governs the relative weight put to L_1 penalization. λ determines the strength of the penalization. A different method to prevent overfitting is the dropout method, proposed by Hinton et al. (2012) and Srivastava et al. (2014). In each iteration of the backpropagation algorithm, a given node is only considered with a probability $1-p$. Consequently, each node is excluded with a probability p which is defined as the dropout rate. The motivation for this is to counteract memorization of the data by preventing co-adaptation of the nodes. Dropout is referred to be an ensemble method, as the final model is a result of training multiple models with reduced capacity.

3 Methodology to Calibrate Systemic Risk

In this section, we explain the details of our systemic risk analysis. Our methodology involves four steps. The first step is concerned with the estimation of VaR based on a linear quantile regression using a set of risk factors as explanatory variables. The results are used in the next step to estimate the CoVaR for each financial institution using a quantile regression neural network. Next, we calculate marginal effects to model systemic risk spillover effects, resulting in a time-varying systemic risk network. In the final step, we propose three systemic risk measures based on this systemic risk network.

Step 1: Estimation of VaR

VaR is defined as the maximum loss over a fixed time horizon at a certain level of confidence. The Basel II Accord introduces VaR as the preferred measure for market risk. The calculation of VaR functions as the basis for capital

requirements of financial institutions. Mathematically, it is the τ-quantile of the return distribution:

$$\mathrm{P}(X_{i,t} \leq \mathrm{VaR}_{i,t}^{\tau}) = \tau, \tag{6}$$

where $X_{i,t}$ is the return of a financial firm i at time t and $\tau \in (0,1)$ is the quantile level. The VaR of each firm i is estimated with a linear quantile regression procedure by regressing the returns on a set of macro state variables M_{t-1}.

$$X_{i,t} = \alpha_i + \gamma_i M_{t-1} + \varepsilon_{i,t}, \tag{7}$$

where the conditional quantile of the error term $Q^{\tau}(\varepsilon_{i,t}|M_{t-1}) = 0$. The VaR is the fitted value of the linear quantile regression problem:

$$\mathrm{VaR}_{i,t}^{\tau} = \widehat{\alpha}_i + \widehat{\gamma}_i M_{t-1}. \tag{8}$$

The validity of a linear model for the estimation of VaR is analyzed by Chao et al. (2015). We refrain from using company-specific balance sheet information due to the low frequency of such data. VaR is a frequently used measure for understanding the critical risk level for an individual financial institution. The drawback of VaR is that it cannot account for determining critical risk levels in a systemic context. Estimating VaR as an individual risk measure is a necessary first step to prepare for calibrating conditional risk.

Step 2: Estimation of CoVaR with Neural Network Quantile Regression

CoVaR was introduced as a systemic extension of standard VaR by Adrian and Brunnermeier (2016). Similar to VaR, it is a risk measure defined as a conditional quantile of the return distribution. But deviating from the VaR concept, CoVaR is contingent on a specific financial distress scenario. The motivation for using CoVaR is the identification of systemically important banks. For the distress scenario, we assume that all other firms are at their VaR. By doing this we follow the reasoning of Hautsch et al. (2014) and Härdle et al. (2016).

$$\mathrm{P}(X_{j,t} \leq \mathrm{CoVaR}_{\mathrm{j,t}}^{\tau} \,|X_{-j,t} = \mathrm{VaR}_{-j,t}^{\tau}) = \tau, \tag{9}$$

where $X_{-j,t}$ is a vector of returns of all firms except j at time t and $\mathrm{VaR}_{-j,t}^{\tau}$ is the corresponding vector of VaRs.

CoVaR can be estimated as a fitted conditional quantile, building on the results for the VaRs obtained in step 1. Chao et al. (2015) and Härdle et al. (2016) find evidence for nonlinearity in the dependence between pairs of financial institutions. Hence, linear quantile regression might not be an appropriate procedure to estimate the risk spillovers, as the interdependencies are potentially different in a state of worsening market conditions. The conditional quantile function of one bank on another may react nonlinearly to the change of critical level of another firm. We therefore propose the use of neural network

quantile regression. The flexibility of the approach allows detecting possible nonlinear dependencies in the data.

The conditional quantile of bank j's returns is regressed on the returns of all other banks and using a neural network as defined in section 2.2:

$$\begin{aligned} X_{j,t} &= h_\theta(X_{-j,t}) + \varepsilon_{j,t}, \\ &= \sum_{m=1}^{M_n} w_m^o \psi \left(\sum_{k \neq j}^{K} w_{k,m}^h X_{k,t} + b_m^h \right) + b^o + \varepsilon_{j,t}, \end{aligned} \tag{10}$$

with the conditional quantile of error term $Q^\tau(\varepsilon_{j,t}|X_{-j,t}) = 0$. To calculate the CoVaR of firm j, the fitted neural network has to be evaluated at the distress scenario:

$$\text{CoVaR}_{j,t}^\tau = \widehat{h}_\theta(\text{VaR}_{-j,t}^\tau), \tag{11}$$

where $\widehat{h}_\theta$ is the estimated neural network. Nonlinearity is introduced by the use of the nonlinear activation function. CoVaR can be interpreted as the hypothetical τ-quantile of the loss distribution if we are in a hypothetical distress scenario. In our case, this distress scenario is all other firms being at their VaR.

Step 3: Calculation of Risk Spillover Effects

Based on the weights estimated by the neural network quantile regression procedure, it is now possible to obtain risk spillover effects between each directed pair of banks. We propose to estimate the spillover effects by taking the partial derivative of the conditional quantile of firm j's return with respect to the return of firm i.

$$\frac{\partial Q^\tau(X_{j,t}|X_{-j,t})}{\partial X_{i,t}} = \frac{\partial}{\partial X_{i,t}} \sum_{m=1}^{M_n} w_m^o \ \psi \left(\sum_{k \neq j}^{K} w_{k,m}^h X_{k,t} + b_m^h \right) + b^o \tag{12}$$

In the case of a sigmoid tangent activation function we have

$$\frac{\partial Q^\tau(X_{j,t}|X_{-j,t})}{\partial X_{i,t}} = \sum_{m=1}^{M_n} w_m^o w_{i,m}^h \psi' \left(\sum_{k \neq j}^{K} w_{k,m}^h X_{k,t} + b_m^h \right) \tag{13}$$

with

$$\psi'(z) = \frac{2}{(\exp^{-z/2} + \exp^{z/2})^2}. \tag{14}$$

In the case of a ReLu activation function we have

$$\frac{\partial Q^\tau(X_{j,t}|X_{-j,t})}{\partial X_{i,t}} = \sum_{m=1}^{M_n} w_m^o w_{i,m}^h \mathbf{I} \left(\sum_{k \neq j}^{K} w_{k,m}^h X_{k,t} + b_m^h > 0 \right), \tag{15}$$

where $\mathbf{I}(\cdot)$ is the indicator function. Note that the non-differentiability of the ReLU function is not an issue in practice since the input of the function is zero with probability zero. As we are interested in the lower tail dependence, we consider the marginal effect evaluated at the distress scenario as defined in the previous subsection:

$$\left.\frac{\partial Q^\tau(X_{j,t}|X_{-j,t})}{\partial X_{i,t}}\right|_{X_{-j,t}=\mathrm{VaR}^\tau_{-j,t}} = \sum_{m=1}^{M_n} w^o_m w^h_{i,m}\ \psi'\left(\sum_{k\neq j}^{K} w^h_{k,m}\,\mathrm{VaR}^\tau_{k,t} + b^h_m\right). \tag{16}$$

Calculating such a marginal effect for each directed pair of firms yields an off-diagonal adjacency matrix of risk spillover effects at time t:

$$A_t = \begin{pmatrix} 0 & a_{12,t} & \dots & a_{1K,t} \\ a_{21,t} & 0 & \dots & a_{2K,t} \\ \vdots & \dots & \ddots & \vdots \\ a_{K1,t} & a_{K2,t} & \dots & 0 \end{pmatrix}, \tag{17}$$

with elements defined as absolute values of marginal effects:

$$a_{ji,t} = \begin{cases} \left|\left.\frac{\partial Q^\tau(X_{j,t}|X_{-j,t})}{\partial X_{i,t}}\right|_{X_{-j,t}=\mathrm{VaR}^\tau_{-j,t}}\right|, & \text{if } j \neq i \\ 0, & \text{if } j = i \end{cases}. \tag{18}$$

Note that the risk spillover effects are not symmetric in general, thus $a_{ji,t} \neq a_{ij,t}$. This adjacency matrix specifies a weighted directed graph modelling the systemic risk in the financial system.

Step 4: Network Analysis of Spillover Effects

To further analyze the systemic relevance of the financial institutions we can calculate several network measures building on the work of Diebold and Yılmaz (2014). They measure the connectedness of financial firms in terms of variance decomposition in a vector autoregressive framework. Their methodology is thus limited to capturing linear spillover effects.

First, the total directional connectedness *to* firm j at time t is defined as the sum of absolute marginal effects of all other firms on j.

$$C_{j\leftarrow\cdot,t} = \sum_{i=1}^{K} a_{ji,t} \tag{19}$$

Analogously, one can define the total directional connectedness *from* firm i at time t as the sum of absolute marginal effects from i to all other firms.

$$C_{\cdot\leftarrow i,t} = \sum_{j=1}^{K} a_{ji,t} \tag{20}$$

Lastly, Diebold and Yılmaz (2014) define the total connectedness at time t as the sum of all absolute marginal effects.

$$C_t = \frac{1}{K} \sum_{i=1}^{K} \sum_{j=1}^{K} a_{ji,t} \tag{21}$$

The total connectedness is a measure for the interconnectedness on the level of the entire system, without differentiating between individual components of the network. Building on this network analysis, we refine the approach by incorporating VaR and CoVaR in the measurement of systemic relevance. In particular, we propose the *Systemic Fragility Index (SFI)* and the *Systemic Hazard Index (SHI)*:

$$SFI_{j,t} = \sum_{i=1}^{K} \left(1 + |\operatorname{VaR}_{i,t}^{\tau}|\right) \cdot a_{ji,t} \tag{22}$$

$$SHI_{i,t} = \sum_{j=1}^{K} \left(1 + |\operatorname{CoVaR}_{j,t}^{\tau}|\right) \cdot a_{ji,t} \tag{23}$$

The *SFI* is a systemic risk measure for the vulnerability of a financial institution. It increases if those adjacency weights pointing to j are large and also if the VaRs of firms i (i.e. the risk factors for j) increase. The *SHI* is a risk measure for the exposure of the financial system to firm i. It depends on the out-going adjacency weights from i and also on the other firms' CoVaRs. As a third measure, we propose the *Systemic Network Risk Index (SNRI)*, a measure for the total systemic risk in the financial system which depends on the marginal effects, the outgoing VaRs, and the incoming CoVaRs.

$$SNRI_t = \sum_{i=1}^{K} \sum_{j=1}^{K} (1 + |\operatorname{VaR}_{i,t}^{\tau}|) \cdot (1 + |\operatorname{CoVaR}_{j,t}^{\tau}|) \cdot a_{ji,t}. \tag{24}$$

Lastly, we define the adjusted adjacency matrix,

$$\widetilde{A}_t = \begin{pmatrix} 0 & \widetilde{a}_{12,t} & \dots & \widetilde{a}_{1K,t} \\ \widetilde{a}_{21,t} & 0 & \dots & \widetilde{a}_{2K,t} \\ \vdots & \dots & \ddots & \vdots \\ \widetilde{a}_{K1,t} & \widetilde{a}_{K2,t} & \dots & 0 \end{pmatrix}, \tag{25}$$

with elements defined as:

$$\widetilde{a}_{ji,t} = \begin{cases} a_{ji,t} \cdot |(1 + |\operatorname{VaR}_{i,t}^{\tau}|) \cdot (1 + |\operatorname{CoVaR}_{j,t}^{\tau}|), & \text{if } j \neq i \\ 0, & \text{if } j = i \end{cases}. \tag{26}$$

The adjusted adjacency matrix accounts for the level of outgoing VaRs and incoming CoVaRs and is an improved representation of risk spillover effects. Systemic spillover effects are thus determined by the marginal effects of the neural network quantile regression procedure as well as by the VaRs and CoVaRs of the considered banks.

4 Empirical Study: US G-SIBs

4.1 Data

For the empirical application of our systemic risk methodology we are focusing on the *global systemically important banks (G-SIBs)* from the United States selected by the Financial Stability Board (FSB), see Table 1. These eight banks constitute systemic risk relevance to the global financial system and are deemed to be too-big-to-fail. We consider daily log returns in a time period between January 4, 2007 and May 31, 2018. The data is obtained from Yahoo Finance.

Financial Institution	NYSE symbol
Wells Fargo & Company	WFC
JP Morgan Chase & co.	JPM
Bank of America Corporation	BAC
Citygroup	C
The Bank of New York Mellon Corporation	BK
State Street Corporation	STT
Goldman Sachs Group, Inc.	GS
Morgan Stanley	MS

Table 1: List of G-SIBs in the USA.

In addition to these stock return data, we consider daily observations of the following set of macro state variables:

i) Implied Volatility Index (VIX), from Yahoo Finance;
ii) the weekly S&P500 index returns, from Yahoo Finance;
iii) Moody's Seasoned Baa Corporate Bond Yield Relative to Yield on 10-Year Treasury Constant Maturity from Federal Reserve Bank of St. Louis;
iv) 10-Year Treasury Constant Maturity Minus 3-Month Treasury Constant Maturity from Federal Reserve Bank of St. Louis.

These macro variables are the common risk factors for the estimation of VaR in the first step of our systemic risk methodology.

4.2 Model Selection

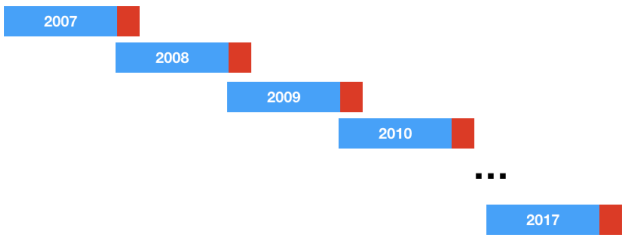


Fig. 1: Rolling window model selection scheme.

The estimation of CoVaR based on neural network quantile regression involves several tuning parameters. Most importantly, we have to make a choice for the activation function and determine the size and structure of the neural network. We select these tuning parameters in a data-driven way. We propose the following model selection procedure.

The data is separated into a training and a validation set repeatedly in a moving window approach. We consider an estimation window of 250 days and a subsequent validation window of 50 days, see Figure 1. A window of 250 observations corresponds to one year of daily return data. For each financial institution, we regress the returns on the other firms' returns to estimate the 5% quantile. The start of each estimation window is the beginning of the new year. The resulting performance indicators are then aggregated over all firms and all windows to select the best model. As a first and most important measure for model performance we propose the average tilted absolute error of prediction ($ATAE$) which is analogous to the MSE in mean regression:

$$ATAE = \frac{1}{n}\sum_{t=1}^{n} \rho_\tau \left\{ X_{j,t} - \widehat{Q}^\tau \left(X_{j,t} | X_{-j,t} \right) \right\}, \tag{27}$$

where $X_{j,t}$ is the observed value and $\widehat{Q}^\tau(X_{j,t}|X_{-j,t})$ is the fitted conditional quantile. A small value for the $ATAE$ is preferred. The second performance measure is the R^1 criterion of Koenker and Machado (1999), which is a coefficient of determination defined analogously to the R^2 measure in mean regression.

$$R^1 = 1 - \frac{\sum_{t=1}^{n} \rho_\tau \left\{ X_{j,t} - \widehat{Q}^\tau \left(X_{j,t} | X_{-j,t} \right) \right\}}{\sum_{t=1}^{n} \rho_\tau \left\{ X_{j,t} - \widehat{Q}^\tau \left(X_{j,t} \right) \right\}}, \tag{28}$$

where $\widehat{Q}^\tau(X_{j,t})$ is the estimated unconditional τ-quantile. The R^1 criterion measures the improvement in model fit compared to the unconditional quantile. It should be noted that for the in-sample fit it has to hold that $R^1 \in [0, 1]$.

However, the out-of-sample fit for a particular unsuitable model can be worse than a constant unconditional quantile fit. In this case, the R^1 can even be negative.

Lastly, we introduce the ratio of quantile exceedances ($RQEX$) as a measure of calibration. A well-calibrated model should have a ratio close to the quantile level τ:

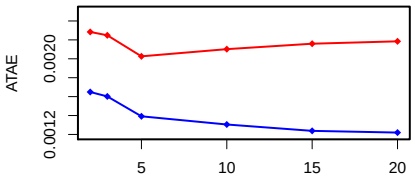
$$RQEX = \frac{1}{n}\sum_{t=1}^{n} \mathbf{I}\left\{X_{j,t} < \widehat{Q}^{\tau}\left(X_{j,t}|X_{-j,t}\right)\right\}. \tag{29}$$

We differentiate between the in-sample and out-of-sample fit. The in-sample fit will always prefer complex models to simple models and will most likely result in overfitting. The out-of-sample fit faces a tradeoff between bias and variance of the prediction and is thus a better measure for generalizing performance. We propose the following model selection scheme based on the out-of-sample performance.

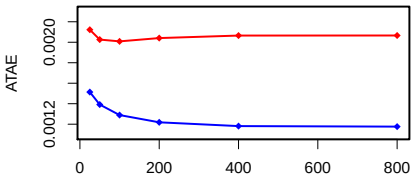
Step 1: Split data in training and test set for each window
Step 2: For each bank j and each window, fit the conditional quantile of X_j contingent on X_{-j} using training data
Step 3: Calculate $ATAE$, R^1 and $RQEX$ for test data
Step 4: Average results over all firms and all windows
Step 5: Choose the model specification with the lowest average $ATAE$

The first tuning parameter we consider is the number of hidden nodes. A large number indicates a high capacity of the model. Figure 2a visualizes the problem of overfitting. Whereas the training error can be effectively reduced, the test error can only be reduced up to a certain point, after which the out-of-sample fit becomes worse. The optimal number of hidden nodes is five. A similar problem occurs with the number of epochs in Figure 2b. A large number of epochs enables the neural network to learn the structure as well as the noise of the training dataset, which makes it prone to overfitting. The optimal number of epochs is 50, conditional on the results of the number of hidden nodes.

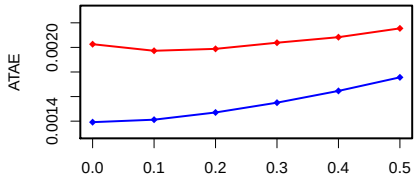
The resulting model structure and complexity of the neural network is rather limited. However, the issue of overfitting should not be disregarded. We therefore consider two regularization methods, namely dropout and elastic net regularization. As expected, both methods have a negative effect on the in-sample fit. However, Figure 2c indicates that a small dropout rate of 10 or 20% leads to better generalizing performance. An explanation for this observation is that by randomly dropping out input variables in each epoch the neural network is prevented from memorizing the data. Dropout is also sometimes referred to as a method of model averaging, as the final model is a result of training several models with reduced complexity. Elastic net, on the other hand, counteracts overfitting by putting an extra penalty on the weight parameters of the neural network. The evidence we find is not conclusive. We check different specifications for λ, governing the strength of the penalty, and



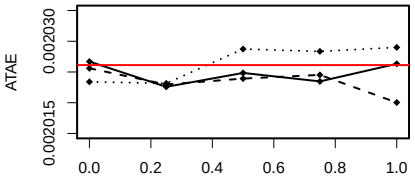
(a) Average $ATAE$ contingent on number of hidden nodes M_n.



(b) Average $ATAE$ contingent on number of epochs.



(c) Average $ATAE$ contingent on input dropout rate p.



(d) Average $ATAE$ contingent on elastic net parameter α. $\lambda = 0.001$ (solid line), $\lambda = 0.005$ (dashed line), $\lambda = 0.01$ (pointed line). The red line indicates the benchmark without elastic net regularization.

Fig. 3: These figures visualize the dependence of the out-of-sample (red line) and in-sample (blue line) average $ATAE$ on different tuning parameters.

Model	$ATAE$	R^1	$RQEX$
ReLU, $M_n = 5$	0.002026	0.4668	0.0650
ReLU, $M_n = 5$, $\alpha = 0.25$, $\lambda = 0.001$	0.002025	0.4676	0.0652
ReLU, $\mathbf{M_n = 5}$, $\mathbf{p = 0.1}$	**0.001973**	**0.4728**	**0.0595**
ReLU, $M_n = 10$	0.002033	0.4788	0.0752
ReLU, $M_n = (5, 5)$	0.002165	0.4513	0.0654
ReLU, $M_n = (10, 3)$	0.002170	0.4397	0.0684
Tanh, $M_n = 2$	0.002845	0.3030	0.1018
Tanh, $M_n = 5$	0.002643	0.3691	0.1036

Table 2: Out-of-sample performance for different model specifications. We consider ReLU and tanh activation functions, M_n refers to the number and structure of hidden nodes, α and λ are the elastic net parameters and p is the input layer dropout rate.

for α, determining the relative weight put on the L_1 and L_2 penalty term. Neither of these specifications leads to significantly improved fit compared to the baseline model with no penalty. The results are illustrated in Figure 2d.

For the final model selection, we consider eight different model specifications. The results can be found in Table 2. We use the ADADELTA optimization algorithm (Zeiler (2012)) with momentum parameter $\rho = 0.99$ and rate annealing $\epsilon = 1e-08$. The number of epochs for all models is 50. The results in Table 2 suggest that complex models are dominated by less complex models. As a second observation, the ReLu (rectifier linear unit) activation function is superior to the tanh activation function. Both dropout and elastic net have a positive impact on the model performance. The best model of the candidates is

a neural network with five hidden nodes in a single hidden layer with a ReLU activation function. The model also has an input dropout ratio of $p = 0.1$. It is ranked first in $ATAE$ (the only model below 0.002) and second in R^1. Also, the model's $RQEX$ (0.0595) is the closest to the theoretical quantile level of 0.05. We will use this model in the following estimation steps.

Finally, we compare the predictive performance of our neural network quantile regression model to a baseline model based on linear quantile regression.

$$X_{j,t} = \beta_0 + \sum_{i \neq j}^{K} X_{i,t}\beta_i + \varepsilon_{j,t}, \tag{30}$$

with $Q^\tau(\varepsilon_t | X_{-j,t}) = 0$. The results stated in Table 3 and 4 clearly show that neural network quantile regression outperforms the linear baseline model. The out-of-sample average $ATAE$ is lower for each window and each firm. This is overwhelming evidence against the linearity hypothesis. The use of a more complex model like a neural network appears to be necessary. A plausible explanation for this is that a linear model is not capable to capture the complex interdependencies of financial firms under distress.

Window	WCF	JPM	BAC	C	BK	STT	GS	MS
1	0.002	0.001	0.002	0.002	0.002	0.002	0.003	0.002
2	0.005	0.005	0.011	0.017	0.006	0.016	0.006	0.007
3	0.001	0.002	0.002	0.003	0.001	0.002	0.001	0.001
4	0.001	0.001	0.002	0.002	0.001	0.001	0.001	0.001
5	0.001	0.001	0.002	0.002	0.001	0.002	0.001	0.001
6	0.001	0.001	0.002	0.001	0.001	0.001	0.001	0.001
7	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
8	0.001	0.001	0.001	0.001	0.001	0.002	0.001	0.001
9	0.001	0.001	0.001	0.001	0.001	0.002	0.001	0.001
10	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
11	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.001

Table 3: Out-of-sample $ATAE$ for final neural network model.

Window	WCF	JPM	BAC	C	BK	STT	GS	MS
1	0.009	0.009	0.008	0.007	0.006	0.006	0.007	0.007
2	0.029	0.022	0.042	0.047	0.014	0.028	0.011	0.014
3	0.007	0.006	0.009	0.008	0.005	0.008	0.007	0.006
4	0.004	0.004	0.005	0.004	0.003	0.003	0.003	0.004
5	0.003	0.005	0.003	0.006	0.003	0.005	0.003	0.004
6	0.001	0.002	0.005	0.005	0.003	0.002	0.003	0.005
7	0.001	0.002	0.002	0.004	0.003	0.004	0.003	0.004
8	0.003	0.004	0.006	0.005	0.004	0.005	0.004	0.004
9	0.006	0.008	0.010	0.011	0.008	0.008	0.009	0.011
10	0.002	0.003	0.005	0.004	0.003	0.004	0.003	0.005
11	0.004	0.004	0.004	0.004	0.003	0.003	0.004	0.004

Table 4: Out-of-sample $ATAE$ for linear model.

4.3 Estimation Results

4.3.1 VaR and CoVaR

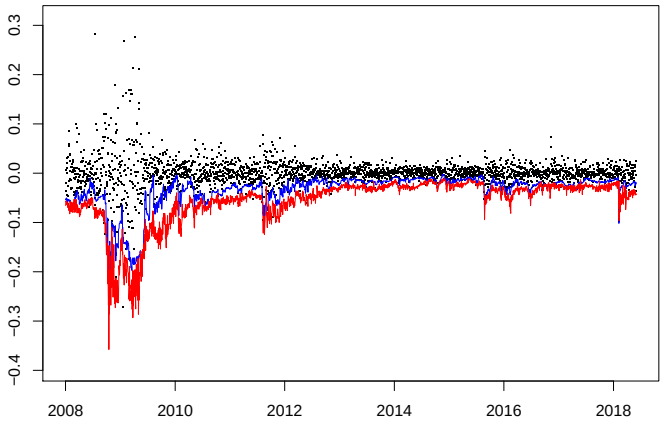


Fig. 4: Plot of Returns (black dots), VaR (blue line) and CoVaR estimated by neural network quantile regression (red line) for Wells Fargo.

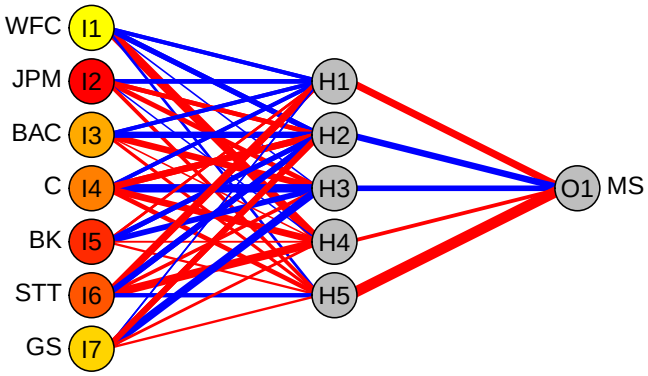


Fig. 5: Fitted quantile regression neural network for Morgan Stanley on March 13, 2008. Red connections indicate negative weights, blue connections indicate positive weights. The color of the input nodes visualizes the variable importance rank calculated as the marginal effect of the respective firm on Morgan Stanley (yellow implies low importance, red implies high importance).

As explained in section 3, the analysis is carried out in four steps. In the first two steps, VaR and CoVaR are estimated for each firm, using linear quantile regression and neural network quantile regression, respectively. To account for potential non-stationarity, we employ a sliding window estimation framework for both measures. The window size is chosen to be 250 observations (representing one year of daily stock returns). We choose a quantile level of $\tau = 5\%$, which is the standard in the related literature, see Hautsch et al. (2014) and Härdle et al. (2016). A lower value for the quantile level leads to less reliable estimates, due to the inverse relation of the variance and the density of the error term. As a sensitivity analysis, we also report the results for $\tau = 1\%$, see Figure 13 and 14 in Appendix B. The results are robust with respect to the choice of the quantile level.

The estimation results for Wells Fargo are visualized in Figure 4. The estimated VaR and CoVaR follow a similar pattern. In the course of the financial crisis and the bankruptcy of Lehman Brothers and Bear Stearns in 2008, both risk measures explode, indicating an increase in systemic risk during this period. A second persistent spike appears in the second half of 2011 caused by the European debt crisis. In the following, both VaR and CoVaR stabilize with a few non-persistent spikes. Similar patterns can be found in the estimation results for the other financial institutions (see Figure 15 in Appendix B). An example of a fitted neural network is visualized in Figure 5.

4.3.2 Risk Spillover Network

Based on the estimation results of the neural network quantile regression procedure and on the fitted VaRs and CoVaRs, we calculate the directional spillover effects for each pair of banks over our prediction horizon. The result is a time-varying weighted adjusted adjacency matrix (as defined in equation 25). This risk spillover network provides insights into the cross-section and the time dynamics of systemic risk. Figure 7 visualizes the evolution of the network in the course of the financial crisis. The first half of 2008 shows a moderate level of lower tail connectedness. This setting changes dramatically in the second half of 2008 with the bankruptcy of Lehman Brothers. As a consequence, the United States Department of the Treasury was compelled to bail out financial institutions to avoid a total collapse of the financial system. Also, the Federal Reserve Bank had to adjust its monetary policy. The time average of the adjacency matrix for 2009 shows a continuing state of financial distress. However, compared to the previous periods one can visually identify a risk cluster in the lower left part of the adjacency matrix. Finally, 2010 shows a decline in systemic risk spillover effects caused by a regained trust in the financial system. Figure 9 restricts the visualization to the largest edges of the financial risk network. As a first observation, spillover effects across banks tend to be symmetric. If bank i has a large impact on bank j, the converse is also very likely. A second observation is the identification of the risk cluster mentioned above. This cluster includes four financial institutions, Citigroup, Bank of America,

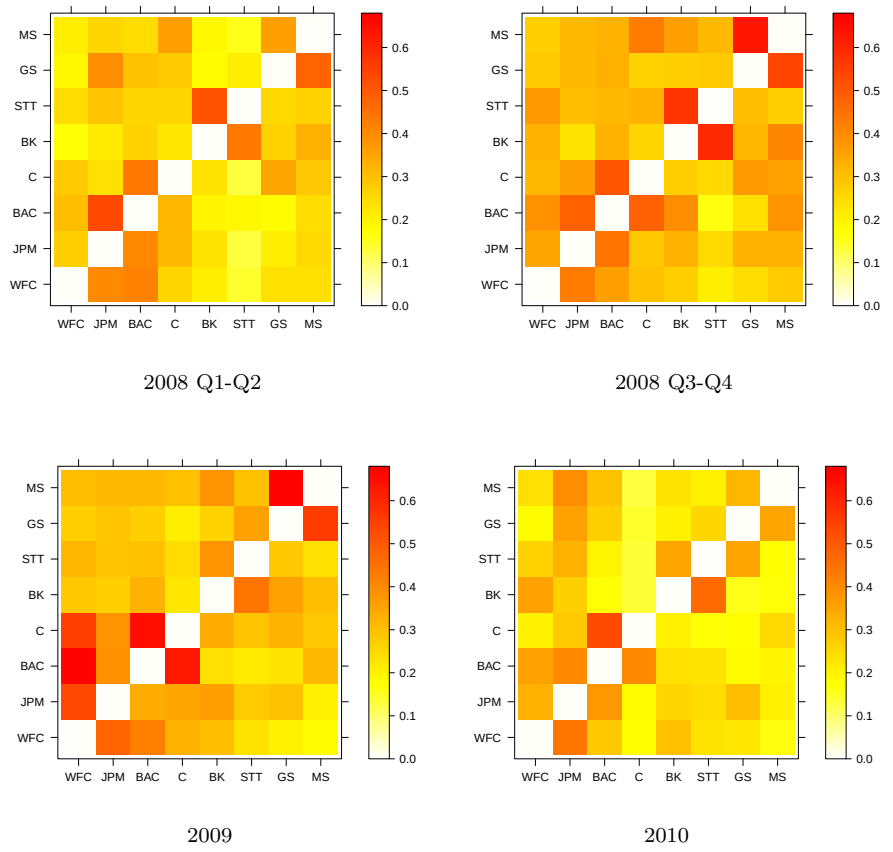


Fig. 7: Time average of risk spillover effects across banks for different time periods.

JP Morgan and Wells Fargo. This cluster coincides with the list of the largest beneficiaries of the bailout program in 2008 and 2009.

4.3.3 Network Risk Measures

Finally, we estimate the systemic risk measures using the results from the previous steps. First, we consider the *Systemic Network Risk Index (SNRI)*, as a measure for total systemic risk in the financial system. Figure 10 shows the development over time. As expected, we see a sharp increase in systemic risk during the financial crisis in the second half of 2008. A second peak appears in the second half of 2011 as a result of the uncertainties associated with the European debt crisis. After a short period of stabilization, we see another rise in systemic risk from 2014 till 2016. In contrast to the previous peaks, this increase appears to be more gradual. When comparing the *SNRI* to the aggregated *SRISK* of Brownlees and Engle (2016), one can identify a co-

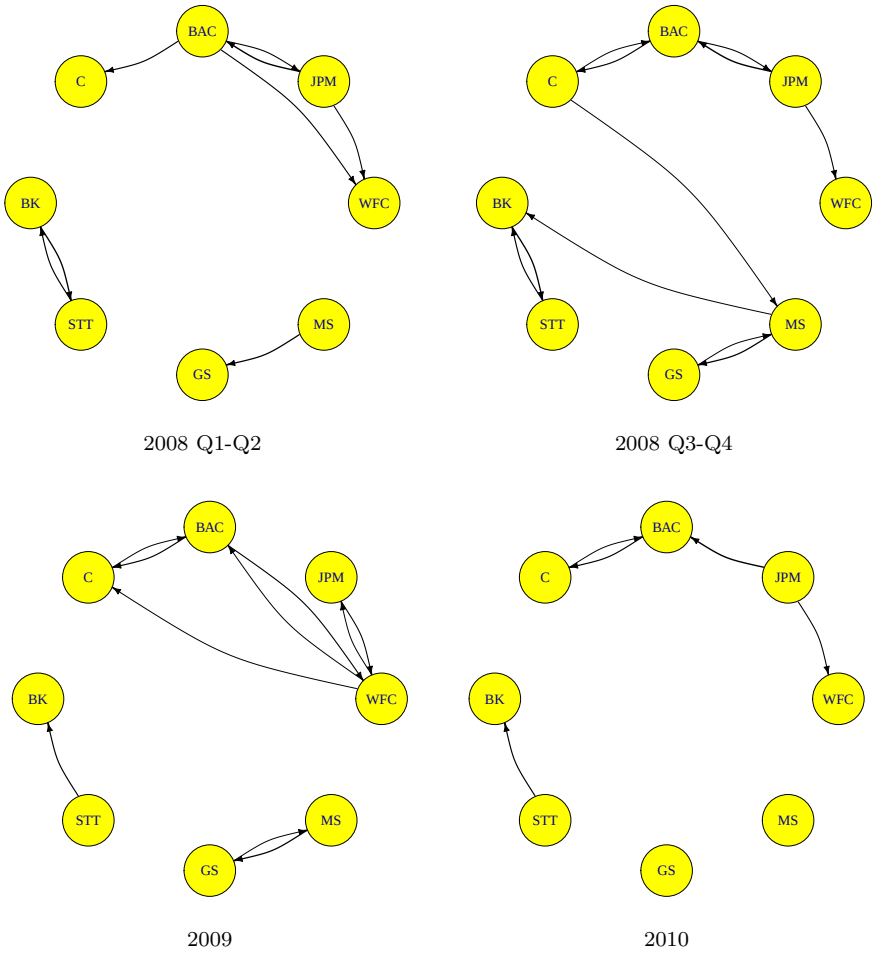


Fig. 9: Time average of risk spillover effects across banks after thresholding ($\widetilde{a}_{ji} > 0.4$) for different time periods.

movement of both indices, see Figure 10. In particular, both the financial crisis and the European debt crisis lead to a sharp increase in both risk measures. However, the aggregated *SRISK* already detects vulnerabilities in the financial system as early as the beginning of 2008. The reason for this is that the $SRISK$ incorporates additional information on micro-prudential variables, namely the book value of debt and the quasi value of assets. In a sense, it is constructed and calibrated to predict the financial crisis "ex-ante", as a high leverage ratio of financial firms proved to be one of the most important factors in hindsight. The second systemically relevant event, namely the European debt crisis, was not primarily caused by high leveraged financial firms. Consequently, we observe

a simultaneous spike of $SNRI$ and aggregated $SRISK$ at the end of 2011. An advantage of the $SNRI$ is that it is entirely based on market data.

While the $SNRI$ is an index for total systemic risk, we also consider firm-specific measures. Table 5 ranks financial firms according to their *Systemic Fragility Index (SFI)*. A large SFI indicates high systemic exposure to the financial system. Our findings suggest that Bank of America and Citigroup are among the most fragile banks during the height of the financial crisis. These results are in line with the findings in Brownlees and Engle (2016). An interesting observation is the ranking of State Street Corporation, which was the first major financial institution to pay back its funds to the US Treasury in July 2009. State Street is ranked on top in 2008 but is among the least fragile banks in 2009 and 2010. Figure 11 shows the time dynamics of the SFI for Citigroup.

We conduct a similar ranking with respect to the *Systemic Hazard Index (SHI)*, which ranks the financial institutions according to the risk they impose on the financial system. In each of the time periods we consider, Bank of America and JP Morgan are listed in the top three. In the aftermath of the crisis in 2009 and 2010 Wells Fargo also emerges as a systemic risk factor to the financial system. Figure 12 visualizes the time dynamics of the SHI for Bank of America. An advantage of our approach is that we are able to differentiate between firms, which affect systemic risk, and firms, which are affected by systemic risk. By doing this we capture the asymmetric nature of systemic risk. As an example, Wells Fargo is ranked high according to the SHI in 2009 and 2010, but at the same time relatively low in SFI. The opposite can be observed for Citigroup, which is ranked low in SHI and high in SFI during the same time periods.

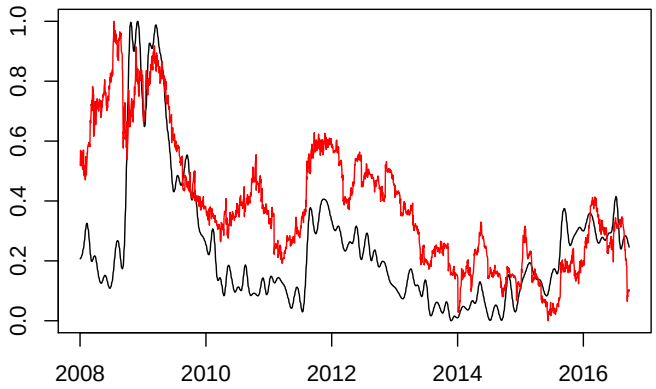


Fig. 10: The figure shows the co-movement of the $SNRI$ (black line) and the $SRISK$ (Brownlees and Engle (2016), red line).

	2008 Q1-Q2		2008 Q3-Q4		2009		2010	
Rank	Ticker	SFI	Ticker	SFI	Ticker	SFI	Ticker	SFI
1	GS	1.998	STT	2.174	C	2.276	BAC	1.883
2	STT	1.871	BK	2.168	BAC	2.219	JPM	1.799
3	BAC	1.861	BAC	2.083	MS	2.180	C	1.738
4	C	1.805	MS	2.035	JPM	2.107	GS	1.727
5	BK	1.804	C	2.029	GS	1.993	BK	1.705
6	WCF	1.741	GS	1,997	BK	1.981	WCF	1.698
7	MS	1.671	JPM	1.962	WCF	1.895	MS	1.696
8	JPM	1.649	WCF	1.849	STT	1.856	STT	1.693

Table 5: The table reports the ranking of financial institutions according to their SFI averaged over different time intervals.

	2008 Q1-Q2		2008 Q3-Q4		2009		2010	
Rank	Ticker	SHI	Ticker	SHI	Ticker	SHI	Ticker	SHI
1	JPM	2.241	BAC	2.317	WCF	2.663	JPM	2.389
2	BAC	2.234	GS	2.239	BAC	2.314	BAC	2.041
3	MS	1.998	JPM	2.204	JPM	2.262	WCF	1.861
4	C	1.906	BK	2.194	GS	2.244	STT	1.738
5	GS	1.789	MS	2.160	BK	2.122	BK	1.730
6	BK	1.679	WCF	2.086	C	2.005	GS	1.653
7	WCF	1.590	C	2.035	STT	1.941	MS	1.447
8	STT	1.341	STT	1.827	MS	1.924	C	1.250

Table 6: The table reports the ranking of financial institutions according to their SHI averaged over different time intervals.

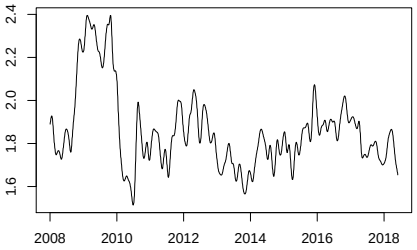


Fig. 11: Time series of the SFI for Citigroup.

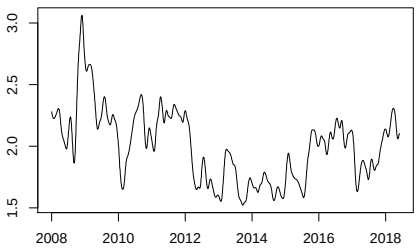


Fig. 12: Time series of the SHI for Bank of America.

5 Conclusion

This paper proposes a novel approach to estimate the conditional value-at-risk (CoVaR) of financial institutions based on neural network quantile regression. Our methodology allows for the identification of risk spillover effects across

banks. We also define three measures for systemic risk, the *Systemic Fragility Index* and the *Systemic Hazard Index* as firm-specific measures and the *Systemic Network Risk Index* as a measure for the overall risk in the financial system. The neural network framework allows us to model systemic risk in a highly nonlinear setting. A comparison to a linear baseline model shows the predictive superiority of our approach in terms of the out-of-sample performance.

We apply our methodology to global systemically important banks from the United States in the period 2007 - 2018. Our results are consistent with previous findings in the literature. We observe the *Systemic Network Risk Index* increasing sharply during the financial crisis and during the European debt crisis. Furthermore, our approach allows to identify a risk cluster of banks which corresponds to the list of banks that receive the largest amount of funding from the US Department of Treasury. By ranking the financial firms according to their *Systemic Fragility Index* and their *Systemic Hazard Index* we are able to identify those firms which bear significant exposure to the financial system and those firms which impose the greatest risk to the financial system.

Appendix A. Consistency of neural network sieve estimator for the conditional quantile

White (1992) shows the consistency of the neural network quantile regression estimator.

Assumption A.1: The data $Z_t = (X_t^\tau, Y_t^\tau)^\tau$ is generated from a bounded stochastic process defined on a complete probability space $(\Omega, \mathcal{F}, P)$, X_t is a random $r \times 1$ vector, Y_t is a random scalar and

(i) Z_t is an i.i.d. process or
(ii) Z_t is a stationary $\phi-$ or $\alpha-$mixing process with such that the mixing coefficients $\phi(k) = \phi_0 \xi^k$ or $\alpha(k) = \alpha_0 \xi^k, 0 < \xi^k < 1, \phi_0, \alpha_0, k > 0$.

Without loss of generality, we may assume $Z_t : \Omega \to \mathbb{I}^{r+1} \stackrel{\text{def}}{=} [0,1]^{r+1}$.

Let $\psi : \mathbb{R} \to \mathbb{R}$ be a bounded function and let (Θ, ρ) be a metric space, where ρ is the L_1-metric. For any $q \in \mathbb{N}$ and $\Delta \in \mathbb{R}^+$ define $T(\psi, q, \Delta) = \{\theta \in \Theta : \theta(x) = \beta_0 + \sum_{j=1}^q \beta_j \psi(x^\top \gamma_j)$ for all x in $\mathbb{I}^r$, $\sum_{j=0}^q |\beta_j| \leq \Delta$, $\sum_{j=1}^q \sum_{i=1}^r |\gamma_{ji}| \leq q\Delta\}$. Further let $Q_n(\theta) = n^{-1} \sum_{t=1}^n |Y_t - \theta(X_t)||\tau - \mathbf{I}(Y_t < \theta(X_t))|$.

Assumption A.2: $\Theta_n(\psi) = T(\psi, q_n, \Delta_n), n = 1, 2, \ldots$, where ψ is bounded, satisfies a Lipschitz condition and is either a cdf or is l-finite. q_n and Δ_n are such that $q_n \to \infty$ and $\Delta_n \to \infty$ as $n \to \infty$. $\Delta_n = o(n^{1/2})$ and either (i) $q_n \Delta_n^2 \log q_n \Delta_n = o(n)$ or (ii) $q_n \Delta_n \log q_n \Delta_n = o(n^{1/2})$.

Assumption A.3: For given quantile level $\tau \in (0,1)$, $\theta_\tau : \mathbb{I}^r \to \mathbb{I}$ is a measurable function such that $\mathrm{P}\{Y_t \leq \theta_\tau(X_t) | X_t\} = \tau$ and for every $\theta \in \Theta$ and all $\epsilon > 0$ sufficiently small $\mathrm{E}\{\theta(X_t) - \theta_\tau(X_t)\} > \epsilon$ implies that for some

$\delta_\epsilon > 0$,

$$\mathrm{E}\left[\mathbf{I}\left\{(\theta_\tau(X_t) + \theta(X_t))/2 \leq Y_t < \theta_\tau(X_t)\right\} | \theta(X_t) < \theta_\tau(X_t)\right] > \delta_\epsilon$$

and

$$\mathrm{E}\left[\mathbf{I}\left\{\theta_\tau(X_t) \leq Y_t < (\theta_\tau(X_t) + \theta(X_t))/2\right\} | \theta(X_t) \geq \theta_\tau(X_t\right] > \delta_\epsilon.$$

Theorem 2.5 White (1992): Given assumptions A.1(i), A.2(i) and A.3 or A.1(ii), A.2(ii) and A.3, there exists a measurable connectionist sieve estimator $\widehat{\theta}_n : \Omega \to \Theta$ such that $Q_n(\widehat{\theta}_n) \leq Q_n(\theta)$, $\theta \in \Theta_n(\psi)$, $n = 1, 2, \ldots$. Further, $\rho(\widehat{\theta}_n, \theta_\tau) \xrightarrow{p} 0$.

Appendix B. Estimation Results

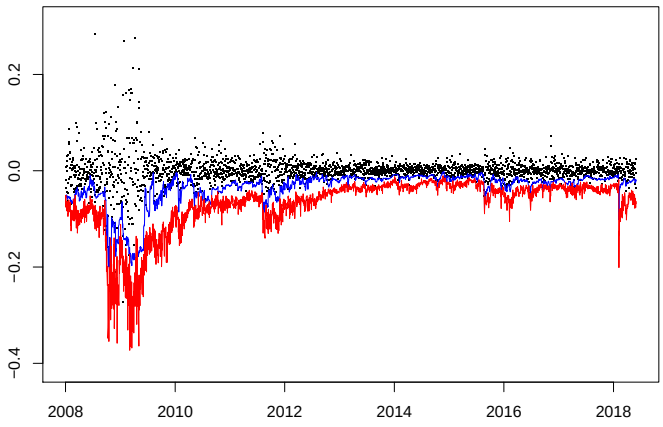


Fig. 13: Plot of Returns (black dots), VaR (blue line) and CoVaR estimated by neural network quantile regression (red line) for Wells Fargo, $\tau = 1\%$.

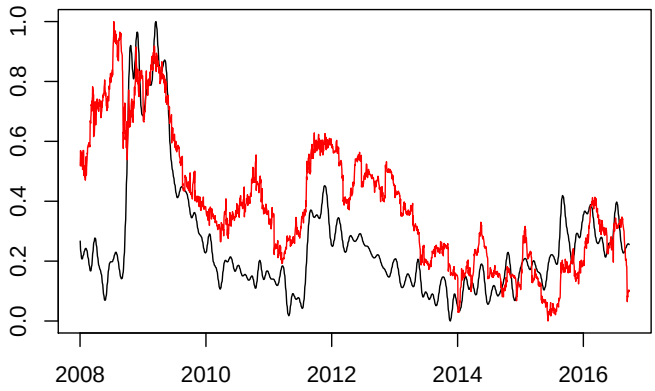


Fig. 14: The figure shows the co-movement of the $SNRI$ (black line) and the $SRISK$ (Brownlees and Engle (2016), red line), $\tau = 1\%$.

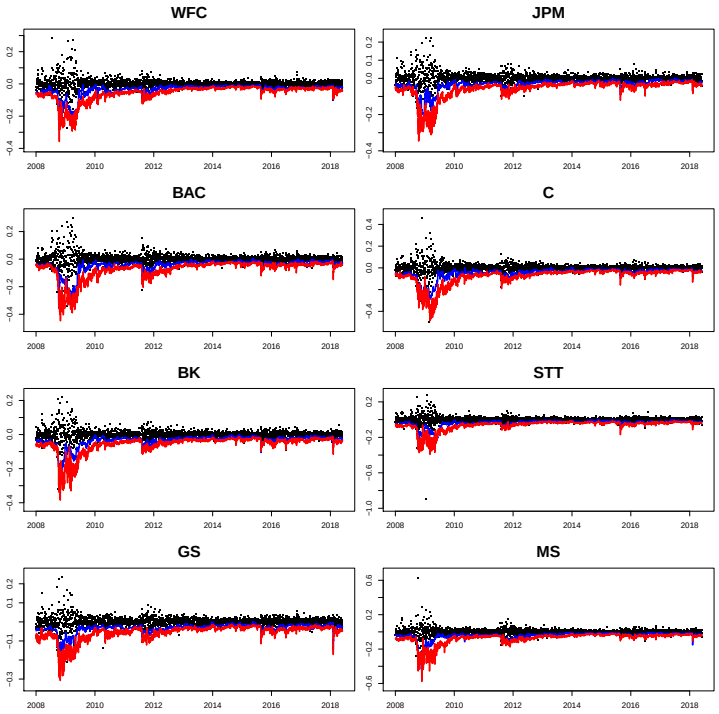


Fig. 15: Plot of Returns (black dots), VaR (blue line) and CoVaR estimated by neural network quantile regression (red line), $\tau = 5\%$.

References

Acharya VV, Pedersen LH, Philippon T, Richardson M (2017) Measuring systemic risk. The Review of Financial Studies 30(1):2–47

Adrian T, Brunnermeier MK (2016) Covar. The American Economic Review 106(7):1705

Brownlees C, Engle RF (2016) Srisk: A conditional capital shortfall measure of systemic risk. The Review of Financial Studies 30(1):48–79

Brownlees CT, Engle R, et al. (2012) Volatility, correlation and tails for systemic risk measurement. Available at SSRN 1611229

Cannon AJ (2011) Quantile regression neural networks: Implementation in r and application to precipitation downscaling. Computers & geosciences 37(9):1277–1284

Chao SK, Härdle WK, Wang W (2015) Quantile regression in risk calibration. Springer

Chen X (2007) Large sample sieve estimation of semi-nonparametric models. Handbook of econometrics 6:5549–5632

Chen X, Shen X (1998) Sieve extremum estimates for weakly dependent data. Econometrica pp 289–314

Chen X, White H (1999) Improved rates and asymptotic normality for nonparametric neural network estimators. IEEE Transactions on Information Theory 45(2):682–691

Cybenko G (1989) Approximation by superpositions of a sigmoidal function. Mathematics of control, signals and systems 2(4):303–314

Diebold FX, Yılmaz K (2014) On the network topology of variance decompositions: Measuring the connectedness of financial firms. Journal of Econometrics 182(1):119–134

Graves A, Mohamed Ar, Hinton G (2013) Speech recognition with deep recurrent neural networks. In: 2013 IEEE international conference on acoustics, speech and signal processing, IEEE, pp 6645–6649

Grenander U (1981) Abstract inference. Tech. rep.

Härdle WK, Wang W, Yu L (2016) Tenet: Tail-event driven network risk. Journal of Econometrics 192(2):499–513

Hautsch N, Schaumburg J, Schienle M (2014) Financial network systemic risk contributions. Review of Finance 19(2):685–738

Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov RR (2012) Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:12070580

Hornik K, Stinchcombe M, White H (1989) Multilayer feedforward networks are universal approximators. Neural networks 2(5):359–366

Johnson J (2018) Deep, skinny neural networks are not universal approximators. arXiv preprint arXiv:181000393

Koenker R, Bassett Jr G (1978) Regression quantiles. Econometrica: journal of the Econometric Society pp 33–50

Koenker R, Bassett Jr G (1982) Robust tests for heteroscedasticity based on regression quantiles. Econometrica: Journal of the Econometric Society pp

43–61

Koenker R, Machado JA (1999) Goodness of fit and related inference processes for quantile regression. Journal of the american statistical association 94(448):1296–1310

Kuan CM, White H (1994) Artificial neural networks: An econometric perspective. Econometric reviews 13(1):1–91

Rolnick D, Tegmark M (2017) The power of deeper networks for expressing natural functions. arXiv preprint arXiv:170505502

Rumelhart DE, Hinton GE, Williams RJ, et al. (1988) Learning representations by back-propagating errors. Cognitive modeling 5(3):1

Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:14091556

Sonoda S, Murata N (2017) Neural network with unbounded activation functions is universal approximator. Applied and Computational Harmonic Analysis 43(2):233–268

Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research 15(1):1929–1958

Taylor JW (2000) A quantile regression neural network approach to estimating the conditional density of multiperiod returns. Journal of Forecasting 19(4):299–311

White H (1988) Economic prediction using neural networks: The case of ibm daily stock returns

White H (1992) Nonparametric estimation of conditional quantiles using neural networks. In: Computing Science and Statistics, Springer, pp 190–199

Xu Q, Liu X, Jiang C, Yu K (2016) Quantile autoregression neural network model with applications to evaluating value at risk. Applied Soft Computing 49:1–12

Zeiler MD (2012) Adadelta: an adaptive learning rate method. arXiv preprint arXiv:12125701

Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. Journal of the royal statistical society: series B (statistical methodology) 67(2):301–320

IRTG 1792 Discussion Paper Series 2019

For a complete list of Discussion Papers published, please visit http://irtg1792.hu-berlin.de.

001 "Cooling Measures and Housing Wealth: Evidence from Singapore" by Wolfgang Karl Härdle, Rainer Schulz, Taojun Xie, January 2019.

002 "Information Arrival, News Sentiment, Volatilities and Jumps of Intraday Returns" by Ya Qian, Jun Tu, Wolfgang Karl Härdle, January 2019.

003 "Estimating low sampling frequency risk measure by high-frequency data" by Niels Wesselhöfft, Wolfgang K. Härdle, January 2019.

004 "Constrained Kelly portfolios under alpha-stable laws" by Niels Wesselhöfft, Wolfgang K. Härdle, January 2019.

005 "Usage Continuance in Software-as-a-Service" by Elias Baumann, Jana Kern, Stefan Lessmann, February 2019.

006 "Adaptive Nonparametric Community Detection" by Larisa Adamyan, Kirill Efimov, Vladimir Spokoiny, February 2019.

007 "Localizing Multivariate CAViaR" by Yegor Klochkov, Wolfgang K. Härdle, Xiu Xu, March 2019.

008 "Forex Exchange Rate Forecasting Using Deep Recurrent Neural Networks" by Alexander J. Dautel, Wolfgang K. Härdle, Stefan Lessmann, Hsin-Vonn Seow, March 2019.

009 "Dynamic Network Perspective of Cryptocurrencies" by Li Guo, Yubo Tao, Wolfgang K. Härdle, April 2019.

010 "Understanding the Role of Housing in Inequality and Social Mobility" by Yang Tang, Xinwen Ni, April 2019.

011 "The role of medical expenses in the saving decision of elderly: a life cycle model" by Xinwen Ni, April 2019.

012 "Voting for Health Insurance Policy: the U.S. versus Europe" by Xinwen Ni, April 2019.

013 "Inference of Break-Points in High-Dimensional Time Series" by Likai Chen, Weining Wang, Wei Biao Wu, May 2019.

014 "Forecasting in Blockchain-based Local Energy Markets" by Michael Kostmann, Wolfgang K. Härdle, June 2019.

015 "Media-expressed tone, Option Characteristics, and Stock Return Predictability" by Cathy Yi-Hsuan Chen, Matthias R. Fengler, Wolfgang K. Härdle, Yanchu Liu, June 2019.

016 "What makes cryptocurrencies special? Investor sentiment and return predictability during the bubble" by Cathy Yi-Hsuan Chen, Roméo Després, Li Guo, Thomas Renault, June 2019.

017 "Portmanteau Test and Simultaneous Inference for Serial Covariances" by Han Xiao, Wei Biao Wu, July 2019.

018 "Phenotypic convergence of cryptocurrencies" by Daniel Traian Pele, Niels Wesselhöfft, Wolfgang K. Härdle, Michalis Kolossiatis, Yannis Yatracos, July 2019.

019 "Modelling Systemic Risk Using Neural Network Quantile Regression" by Georg Keilbar, Weining Wang, July 2019.

IRTG 1792, Spandauer Strasse 1, D-10178 Berlin
http://irtg1792.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the IRTG 1792.